IN THE CLAIMS

This listing of claims replaces all prior listings:

1. (currently amended) A method for balancing a load in a network having a load balancing slave server, a load balancing master server, a plurality of processing servers, and a client, the method comprising the steps of:

assigning by the load balancing master server the load balancing slave server to receive a request from the client to perform a processing;

receiving at the load balancing slave server the request from the client to perform the processing;

sending by the load balancing slave server the request to the load balancing master server in response to the receipt of the request;

determining a load of each of the plurality of processing servers by the load balancing master server;

selecting by the load balancing master server a selected one of the plurality of processing servers that is suitable for performing the processing, wherein the selected one of the plurality of processing servers is selected based on the load of each of the plurality of processing servers;

sending an identifier of the selected one of the plurality of processing servers from the load balancing master server to the load balancing slave server; and

establishing by the load balancing slave server a communication link between the selected one of the plurality of processing servers and the client to perform the processing.

2. (currently amended) The method of claim 1, wherein the step of establishing further includes the step of:

routing the communication link between the selected one of the plurality of processing servers and the client through the load balancing slave server.

3. (currently amended) The method of claim 1, further comprising the step of:

receiving a plurality of load metrics from each of the plurality of processing servers.

4. (currently amended) The method of claim 1, wherein the step of determining further comprises the step of:

receiving a load metric with the request from the load balancing slave server at the load balancing master server.

5. (currently amended) A method in a data processing system having a first and a second load balancing server and having a plurality of processing servers, the method comprising the steps of:

assigning by the second load balancing server the first load balancing server to receive a request from a client to perform a processing;

receiving by the first load balancing server the request to perform the processing;

sending the request from the first load balancing server to the second load balancing server;

determining a load of each of the plurality of processing servers by the second load balancing server;

selecting by the second load balancing server a selected one of the plurality of processing servers that is suitable for performing the processing, wherein the selection is performed based on the load of each of the plurality of processing servers;

sending an identifier of the selected one of the plurality of processing servers from the second load balancing server to the first load balancing server; and

sending by the second load balancing server to the selected one of the plurality of processing servers an indication to perform the processing.

6. (previously presented) The method of claim 5, wherein the step of sending by the second load balancing server further comprises the step of:

identifying to the first load balancing server the selected one of the plurality of processing servers after the indication to perform the processing has been sent to the selected one of the plurality of processing servers.

7. (original) The method of claim 5, further comprising the steps of:

receiving a plurality of load metrics that originate from the plurality of processing servers at the second load balancing server.

8. (previously presented) The method of claim 5, wherein sending the request further includes the step of:

encoding at least one load metric in the request.

9. (previously presented) The method of claim 5, wherein the first load balancing server is a load balancing slave.

10. (previously presented) The method of claim 5, wherein the second load balancing server is a load balancing master.

11. (currently amended) A data processing system, comprising:

a plurality of processing servers;

a client sends a request;

a load balancing slave server that is assigned by a load balancing master server to receive the request from the client, that sends the request to an external source for a selection of one of the plurality of processing servers that is suitable for performing a processing, that receives an indication of the selected one of the plurality of processing servers from the external source, and that establishes a communication link between the selected one of the plurality of processing servers and the client to perform the processing; and

the load balancing master server that assigns the load balancing slave server to receive the request from the client, that receives the request from the load balancing slave server, that determines a load of each of the plurality of processing servers, that selects the selected one of the plurality of processing servers based on the load of each of the plurality of processing servers, and that sends the indication of the selected one of the plurality of processing servers to the load balancing slave server.

12. (currently amended) The data processing system of claim 11, wherein a plurality of load metrics are received at the load balancing master server from the plurality of processing servers that indicate the load on each of the plurality of processing servers.

13. (currently amended) The data processing system of claim 11, wherein at least one load metric is included in the request sent by the load balancing slave server to the external source.

14. (previously presented) A data processing system, comprising:

a plurality of processing servers;

a client that sends a request to have processing performed in a load balanced manner;

a first load balancing server that is assigned by a second load balancing server to receive the request from the client and that receives the request from the client; and

the second load balancing server that assigns the first load balancing server to receive the request from the client, that receives the request from the first load balancing server, that determines a load of each of the plurality of processing servers, that selects a selected one of the plurality of processing servers that is suitable for performing the processing in the load balanced manner, and that sends to the selected one of the plurality of processing servers an indication to perform the processing, wherein the selection is based on the load of each of the plurality of processing servers.

15. (original) The data processing system of claim 14, wherein the first load balancing server is a load balancing slave.

16. (original) The data processing system of claim 14, wherein the second load balancing server is a load balancing master.

17. (previously presented) The data processing system of claim 14, wherein the second load balancing server is in receipt of a plurality of load metrics that originate from each of the plurality of processing servers and indicate the load on each of the plurality of processing servers.

18. (currently amended) A computer-readable medium containing instructions that cause a data processing system to perform a method for balancing a load in a network having a load balancing slave server, a load balancing master server, a plurality of processing servers, and a client, the method comprising the steps of:

assigning by the load balancing master server the load balancing slave server to receive a request from the client to perform a processing;

receiving at the load balancing slave server the request from the client to perform the processing;

sending by the load balancing slave server the request to the load balancing master server in response to the receipt of the request;

determining a load of each of the plurality of processing servers by the load balancing master server;

selecting by the load balancing master server a selected one of the plurality of processing servers that is suitable for performing the processing, wherein the selected one of the plurality of processing servers is selected based on the load of each of the plurality of processing servers;

sending an identifier of the selected one of the plurality of processing servers from the load balancing master server to the load balancing slave server; and

establishing by the load balancing slave server a communication link between the selected one of the plurality of processing servers and the client to perform the processing.

19. (currently amended) The computer-readable medium of claim 18, wherein the step of establishing further includes the step of:

routing the communication link between the selected one of the plurality of processing servers and the client through the load balancing slave server.

20. (currently amended) The computer readable medium of claim 18, further comprising the step of:

receiving a plurality of load metrics from each of the plurality of processing servers.

21. (currently amended) The computer readable medium of claim 18, wherein the step of determining further comprises the step of:

receiving a load metric with the request from the load balancing slave server at the load balancing master server.

22. (previously presented) A computer readable medium containing instructions that cause a data processing system to perform a method for load balancing having a first and a second load balancing server and having a plurality of processing servers, the method comprising the steps of:

assigning by the second load balancing server the first load balancing server to receive a request from a client to perform a processing;

receiving by the first load balancing server the request to perform the processing;

sending the request from the first load balancing server to the second load balancing server;

determining a load of each of the plurality of processing servers by the second load balancing server;

selecting by the second load balancing server a selected one of the plurality of processing servers that is suitable for performing the processing, wherein the selection is performed based on the load of each of the plurality of processing servers; and

sending by the second load balancing server to the selected one of the plurality of processing servers an indication to perform the processing.

23. (previously presented) The computer-readable medium of claim 22, wherein the step of sending by the second load balancing server further comprises the step of:

identifying to the first load balancing server the selected one of the plurality of processing servers after the indication to perform the processing has been sent to the selected one of the plurality of processing servers.

24. (original) The computer-readable medium of claim 22, further comprising the steps of:

receiving a plurality of load metrics that originate from the plurality of processing servers at the second load balancing server.

25. (previously presented) The computer-readable medium of claim 22, wherein sending the request further includes the step of:

encoding at least one load metric in the request.

26. (currently amended) A load balancer for balancing a load in a network having a load balancing slave server, a load balancing master server, a plurality of processing servers, and a client, the method comprising the steps of:

means for assigning by the load balancing master server the load balancing slave server to receive a request from a client to perform a processing;

means for receiving at the load balancing slave server the request from the client to perform the processing;

means sending by the load balancing slave server the request to the load balancing master server in response to the receipt of the request;

means for determining a load of each of the plurality of processing servers by the load balancing master server;

means for selecting by the load balancing master server a selected one of the plurality of processing servers that is suitable for performing the processing, wherein the selected one of the plurality of processing servers is selected based on the load of each of the plurality of processing servers;

means for sending an identifier of the selected one of the plurality of processing servers from the load balancing master server to the load balancing slave server; and

means for establishing by the load balancing slave server a communication link between the selected one of the plurality of processing servers and the client to perform the processing.